Dependence Modelling Across Major Causes of Death via Time-Varying Copula State Space Models

Ariane Hanebeck

Scientific Computing Center, Karlsruhe Institute of Technology

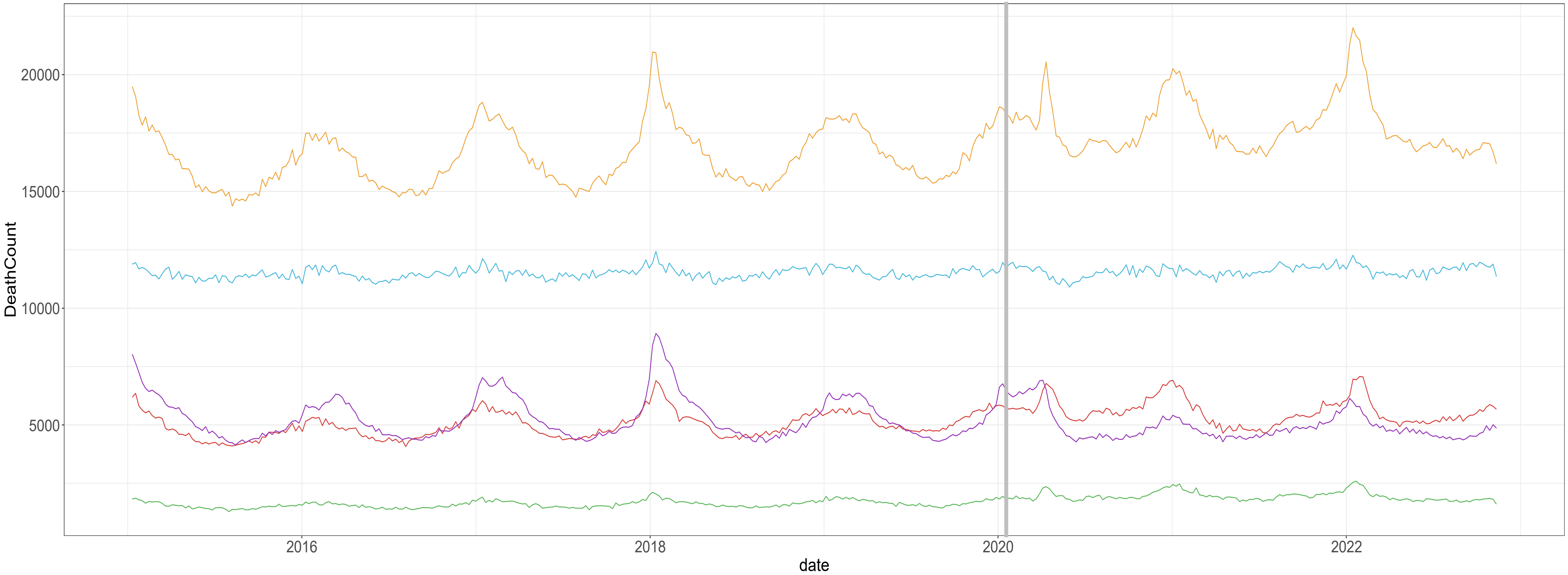
Collaborators:
Dr. Han Li, Associate Professor, The University of Melbourne
Prof. Ph.D. Claudia Czado, Technical University of Munich

South-West German Infectious Disease Modelling Workshop 2025

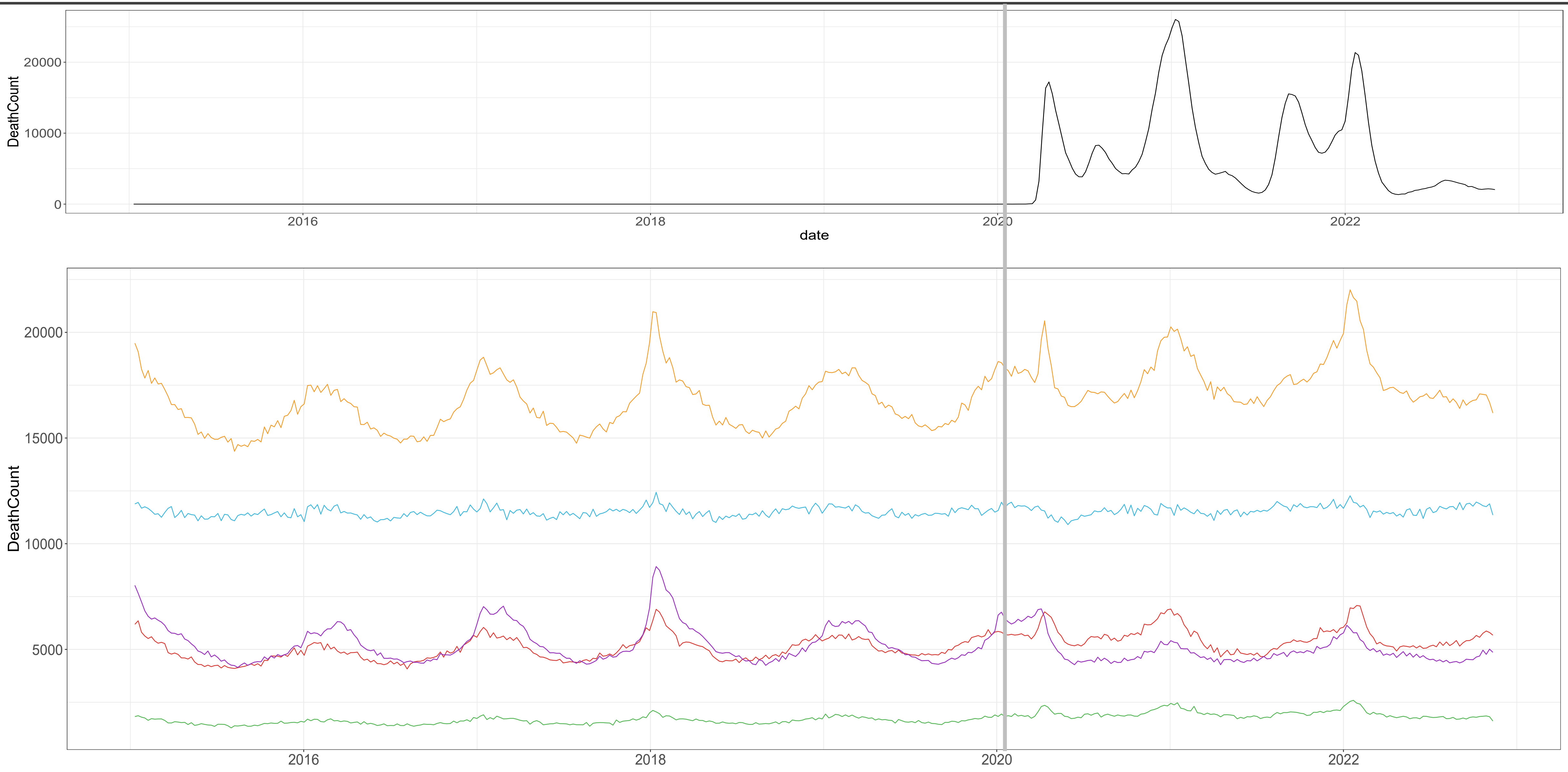
OUTLINE

DATA DESCRIPTION

- US weekly data in time frame January 2015 – November 2022
- Most common conditions reported on death certificates where COVID-19 was listed
- Death counts of Circulatory, Alzheimer, Respiratory, Diabetes, Malignant



DATA DESCRIPTION



MOTIVATION

- Data shows clear changes between periods before and after COVID-19 started

- Motivation: What is the influence of COVID-19 on major death counts?
  1. Marginal structure: Assess level changes in mortality before/after COVID-19 started
  2. Dependence structure: Understand dependence structure among causes and whether it changes with COVID-19

→ Separate marginal behavior from the depencence structure

→ Copula approach

OUTLINE

NOTATION

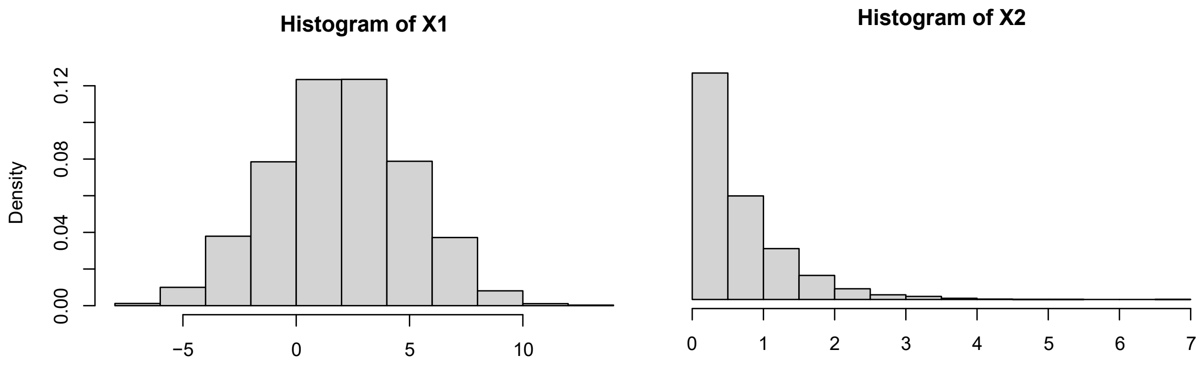
- For this talk: Only interested in bivariate copulas

- Notation:
  - Consider random variables X_1, X_2 with
  - marginal distribution functions F_1, F_2,
  - marginal densities f_1, f_2, and
  - joint distribution function F

- Example: $X_1 \sim F_1, X_2 \sim F_2$ independent

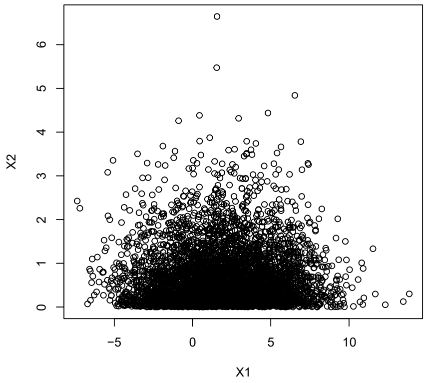
INTRODUCTION COPULAS (I)

$X_1 \sim F_1, X_2 \sim F_2$ independent



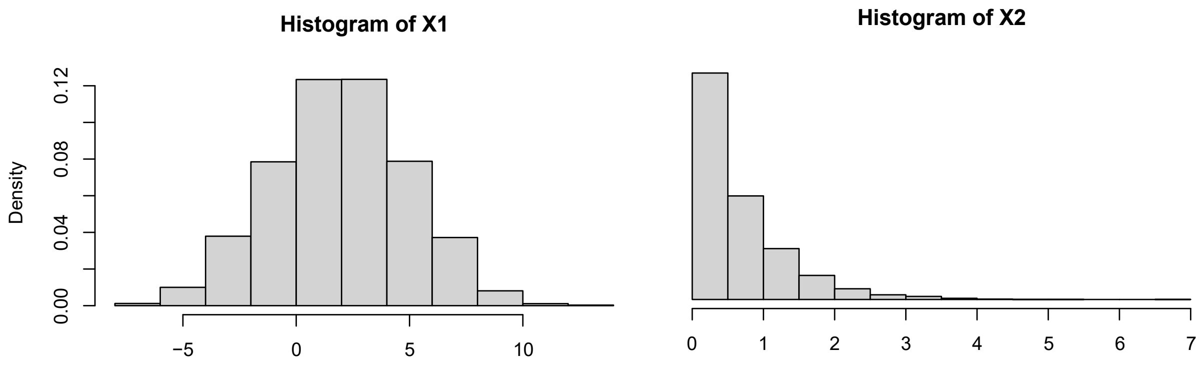
$X_1 \sim N(2{,}3), X_2 \sim \mathrm{Exp}(1.5)$

X-Scale: X_i



INTRODUCTION COPULAS (I)

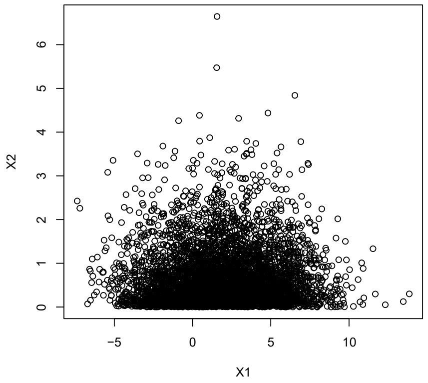
$X_1 \sim F_1, X_2 \sim F_2$ independent



$X_1 \sim N(2{,}3), X_2 \sim \mathrm{Exp}(1.5)$

- PIT: $X_i \sim F_i \Rightarrow U_i = F_i(X_i) \sim U[0{,}1]$
- Inverse PIT: $U_i \sim U[0{,}1] \Rightarrow F_i^{-1}(U_i) \sim F_i$

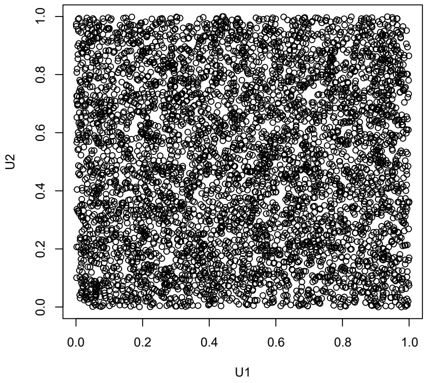
X-Scale: X_i



$$U_i = F_i(X_i)$$

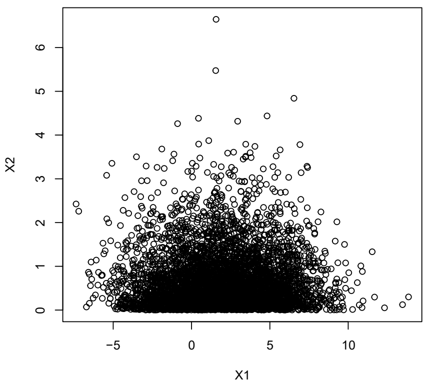
$$X_i = F_i^{-1}(U_i)$$

U-Scale: U_i



INTRODUCTION COPULAS (I)

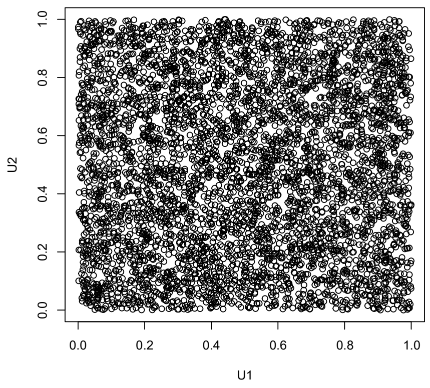
X-Scale: X_i



$$U_i = F_i(X_i)$$

$$X_i = F_i^{-1}(U_i)$$

U-Scale: U_i

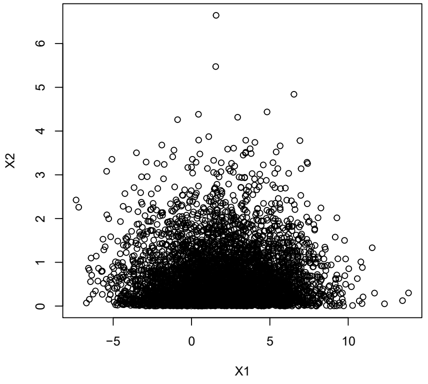


- Copulas separate margins from dependence structure
- Bivariate copula C: distribution function on $[0,1]^2$ with uniformly distributed margins
- Sklar‘s Theorem (1959):

$$F(x_1, x_2) = C\big(F_1(x_1), F_2(x_2)\big)$$

INTRODUCTION COPULAS (I)

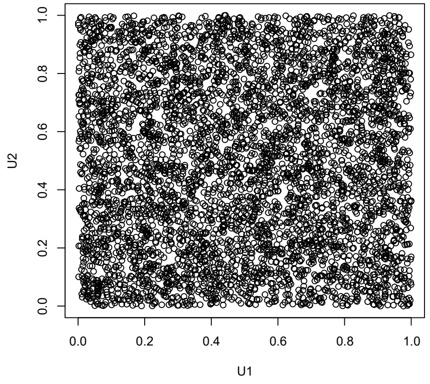
X-Scale: X_i



$$U_i = F_i(X_i)$$

$$X_i = F_i^{-1}(U_i)$$

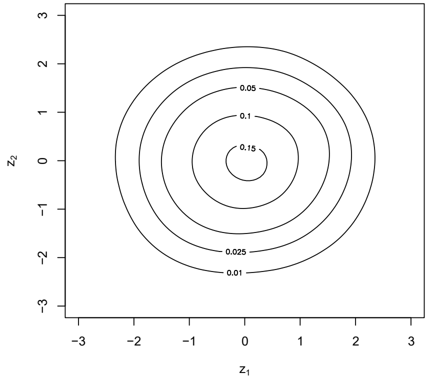
U-Scale: U_i



$$Z_j = \Phi^{-1}(U_j)$$

$$U_j = \Phi(Z_j)$$

Z-Scale: Z_i



- Copulas separate margins from dependence structure
- Bivariate copula C: distribution function on $[0,1]^2$ with uniformly distributed margins
- Sklar‘s Theorem (1959):

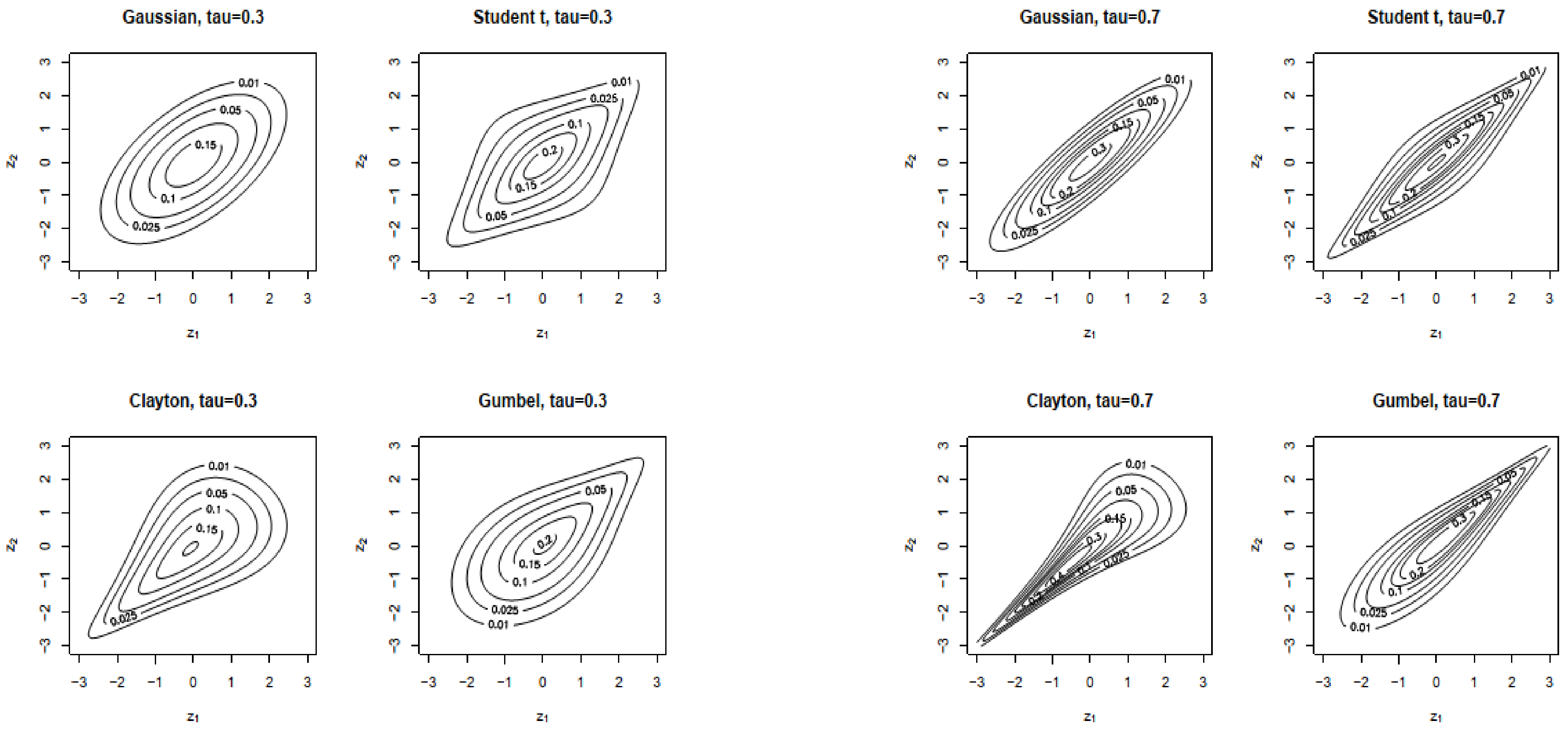
$$F(x_1, x_2) = C\big(F_1(x_1), F_2(x_2)\big)$$

Use contour plots on z-scale

- Identify signature shapes of common bivariate copula families
- Here: Circles correspond to independence

INTRODUCTION COPULAS (2)

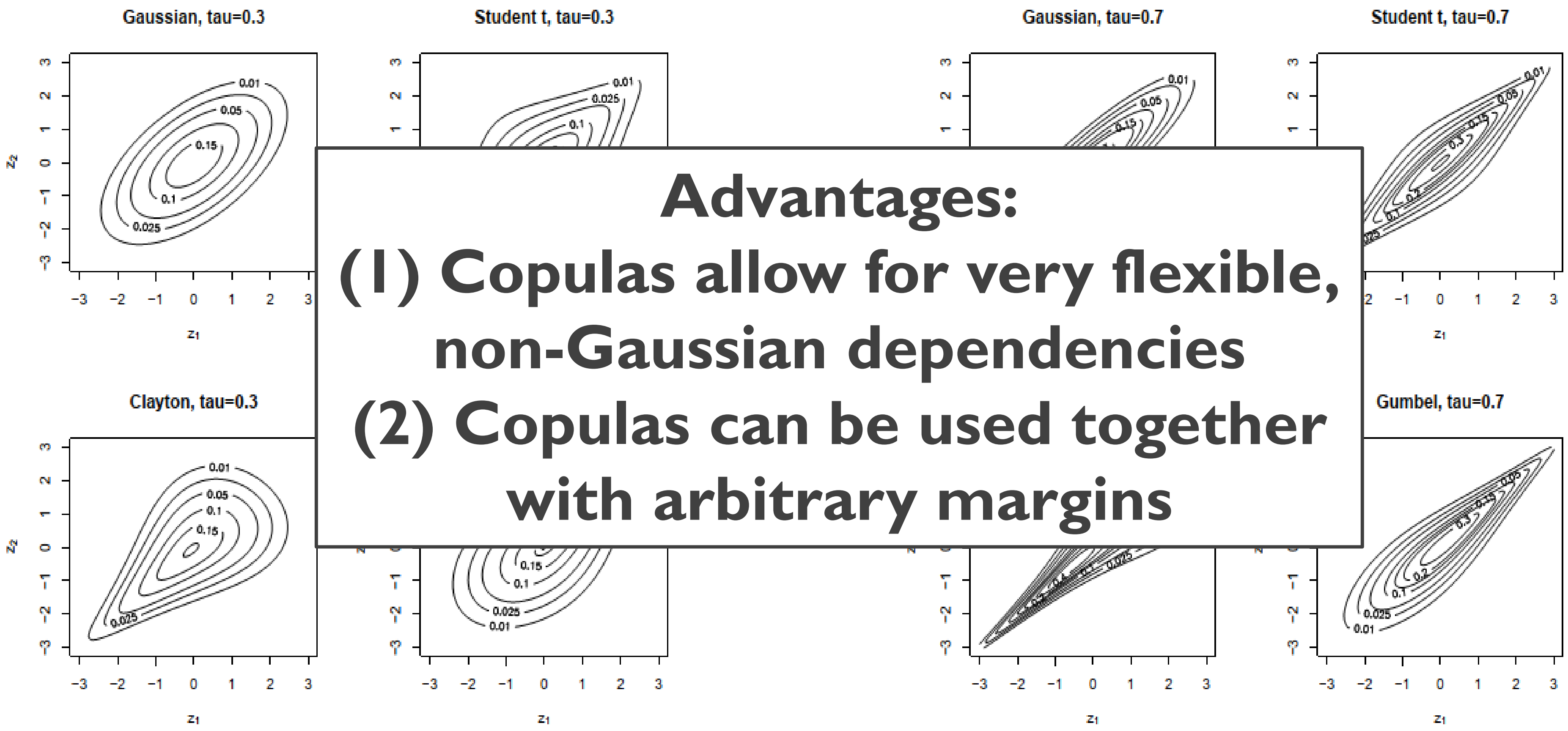
Examples of bivariate copula families:



Kendall's τ: Dependence measure in [-1,1], does not depend on marginal distributions

INTRODUCTION COPULAS (2)

Examples of bivariate copula families:



Advantages:

(1) Copulas allow for very flexible, non-Gaussian dependencies

(2) Copulas can be used together with arbitrary margins

Kendall‘s τ: Dependence measure in [-1,1], does not depend on marginal distributions

OUTLINE

MARGINAL MODELING

Original Scale

Start with mortality data on the original scale: $Y_{t,j}, t = 1, \ldots, T, j = 1, \ldots, 5$

1 = Circulatory
2 = Alzheimer
3 = Respiratory
4 = Diabetes
5 = Malignant

MARGINAL MODELING

Original Scale

Start with mortality data on the original scale: $Y_{t,j}, t = 1, \ldots, T, j = 1, \ldots, 5$

Decompose

Decompose each cause $Y_{t,j}$ into Trend $T_{t,j}$, Seasonality $S_{t,j}$, Covid Deaths CD_t, Indicator I_t, Remainder $R_{t,j}$

$$Y_{t,j} = T_{t,j} + S_{t,j} + \delta_{1,t} \cdot CD_t + \delta_2 \cdot I_t + R_{t,j}$$

I_t is 1 before COVID-19 started and 0 afterward

Decomposition using [Dokumentov and Hyndman 2022]

MARGINAL MODELING

Original time series

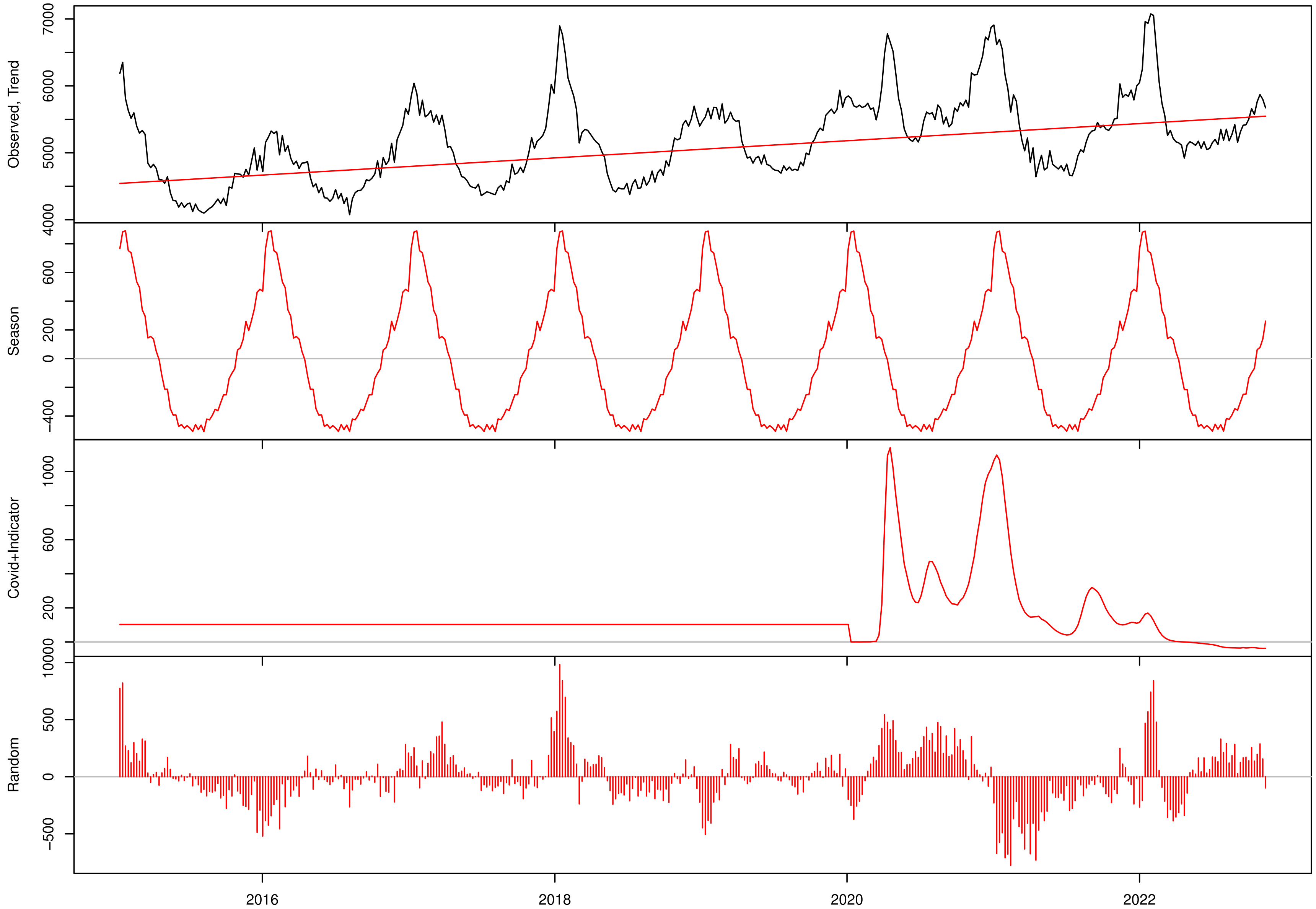
Trend $\hat{T}_{t,j}$

Seasonality $\hat{S}_{t,j}$

COVID-19 and Indicator

$\hat{\delta}_{1,t} \cdot CD_t + \hat{\delta}_2 \cdot I_t$

Remainer $\hat{R}_{t,j}$



MARGINAL MODELING

Original Scale

Start with mortality data on the original scale: $Y_{t,j}, t = 1, \dots, T, j = 1, \dots, 5$

Decompose

Decompose each cause $Y_{t,j} = T_{t,j} + S_{t,j} + \delta_{1,t} \cdot CD_t + \delta_2 \cdot I_t + R_{t,j}$

Residual Scale

Obtain standardized residuals

$$\hat{\epsilon}_{t,j} = \frac{y_{t,j} - \hat{y}_{t,j}}{\hat{\sigma}_j}$$

MARGINAL MODELING

Original Scale

Start with mortality data on the original scale: $Y_{t,j}, t = 1, \ldots, T, j = 1, \ldots, 5$

Decompose

Decompose each cause $Y_{t,j} = T_{t,j} + S_{t,j} + \delta_{1,t} \cdot CD_t + \delta_2 \cdot I_t + R_{t,j}$

Residual Scale

Obtain standardized residuals

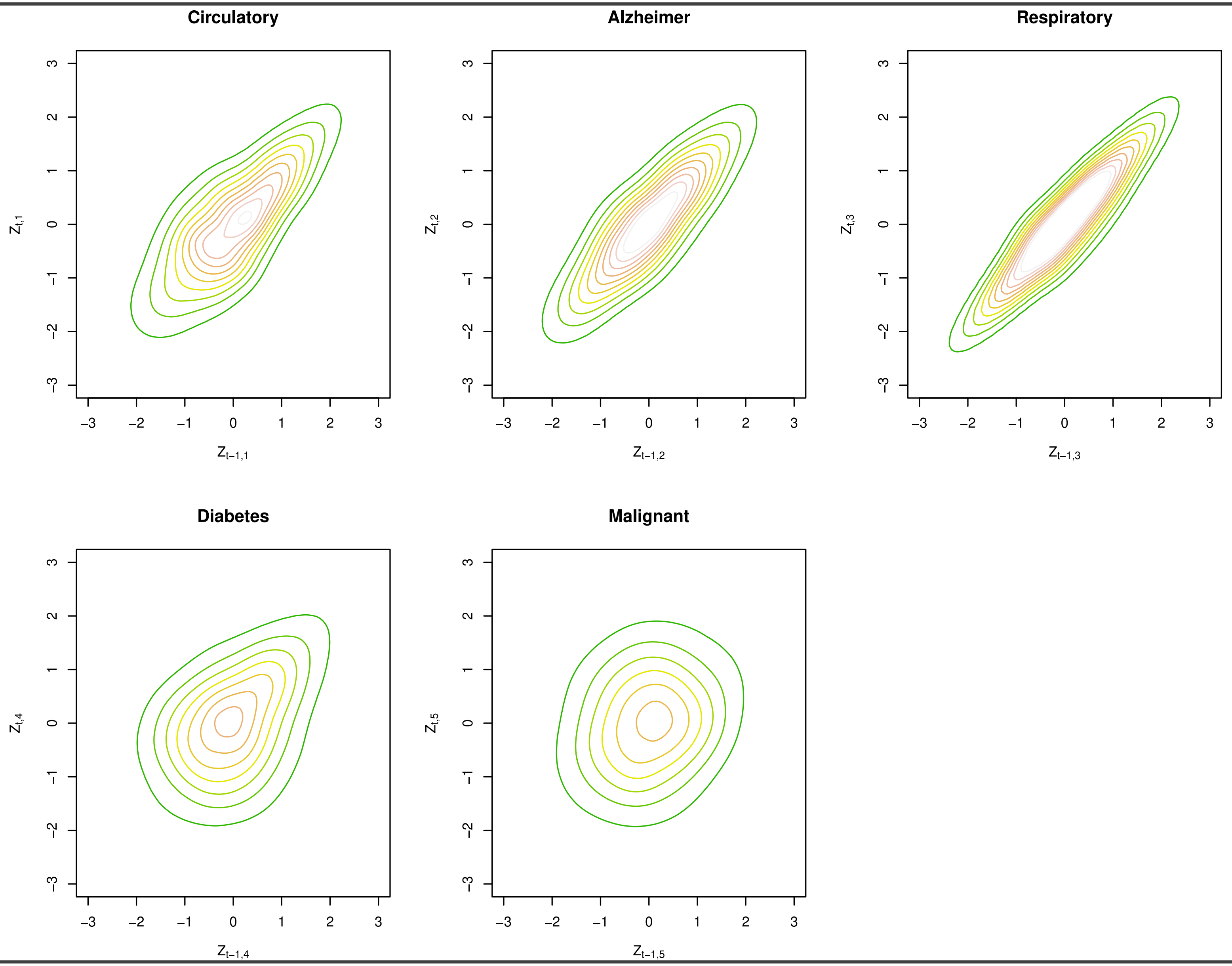
$$\hat{\epsilon}_{t,j} = \frac{y_{t,j} - \hat{y}_{t,j}}{\hat{\sigma}_j}$$

U-Scale

Pseudo copula data can be obtained by $\hat{u}_{t,j} = \hat{F}_j(\hat{\epsilon}_{t,j})$

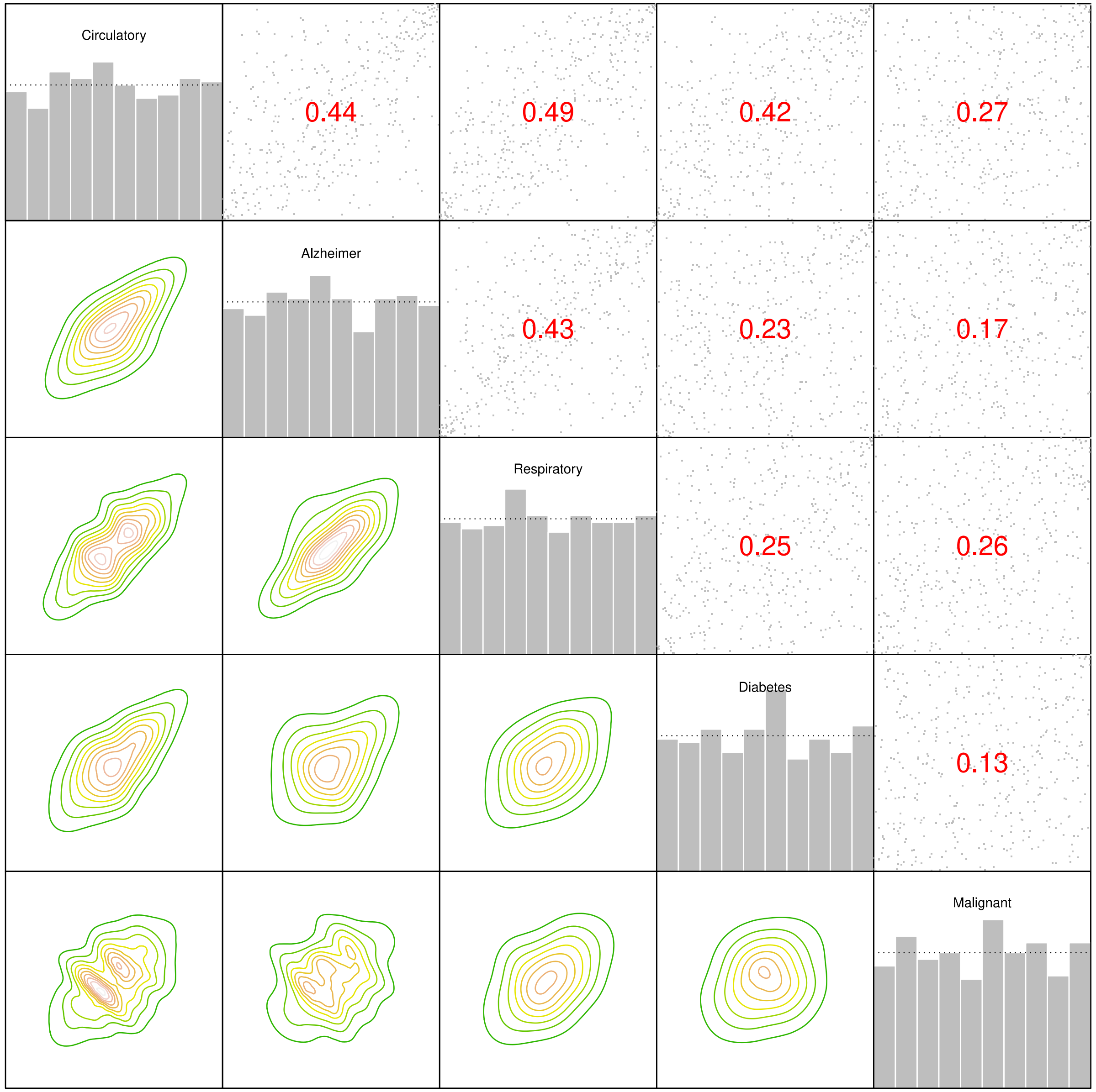
OUTLINE

1. PROPERTY: SERIAL DEPENDENCE



- Consider contour plots of
  $$\left(\hat{z}_{t,j}, \hat{z}_{t-1,j}\right) = \left(\Phi^{-1}\left(\hat{u}_{t,j}\right), \Phi^{-1}\left(\hat{u}_{t-1,j}\right)\right)$$
- Strong non-Gaussian serial dependence (except malignant)

2. PROPERTY: CROSS-SECTIONAL DEPENDENCE

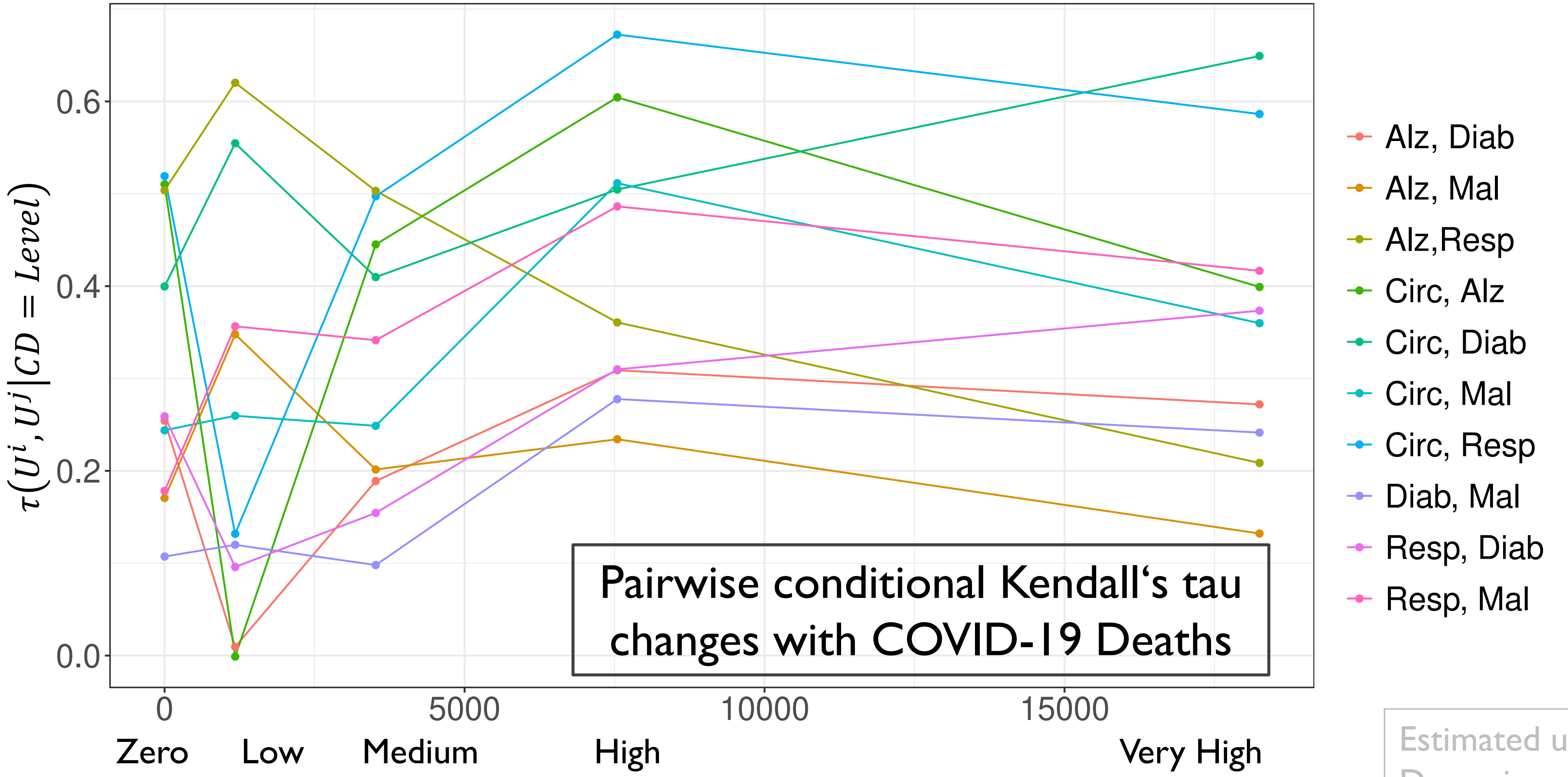


- Upper triangle: u-data scatter and Kendall‘s tau
- Diagonal: Histograms of u-data
- Lower triangle: Contours of

$$\left(\hat{z}_{t,j}, \hat{z}_{t,k}\right) = \left(\Phi^{-1}\left(\hat{u}_{t,j}\right), \Phi^{-1}\left(\hat{u}_{t,k}\right)\right)$$

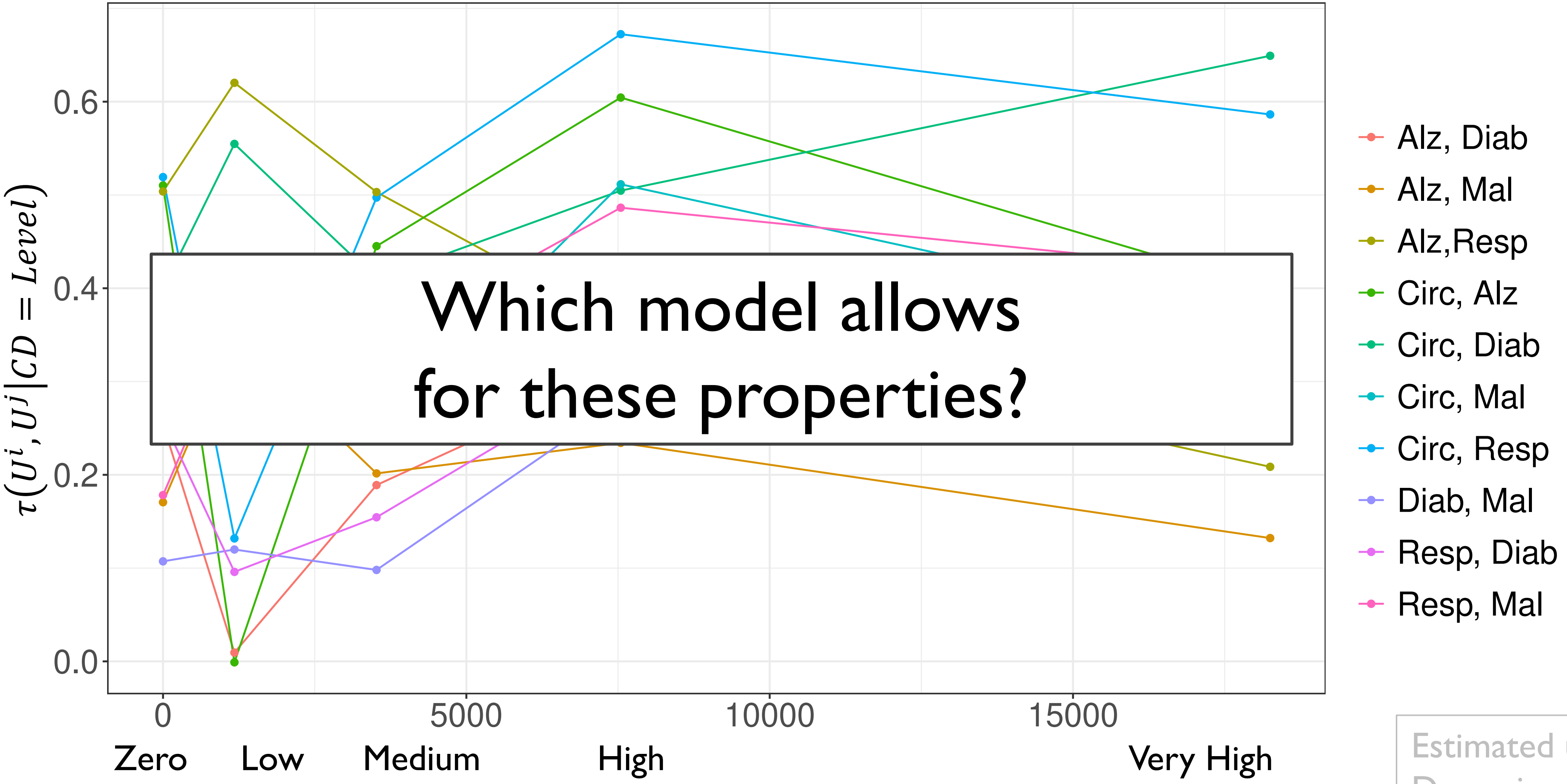
- Non-Gaussian cross-sectional dependence visible

3. PROPERTY: KENDALL'S TAU



Estimated using Derumigny (2019)

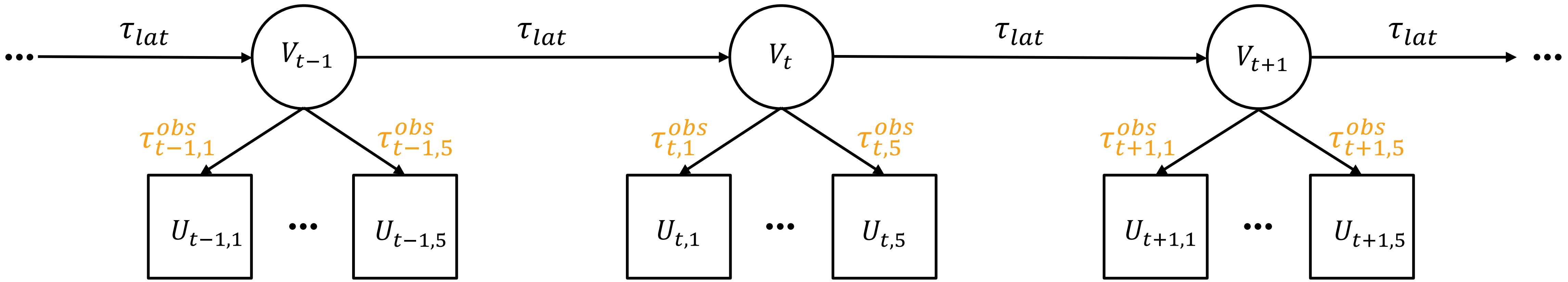
3. PROPERTY: KENDALL'S TAU



Which model allows for these properties?

Estimated using Derumigny (2019)

COPULA-SSM WITH TIME-VARYING PARAMETERS

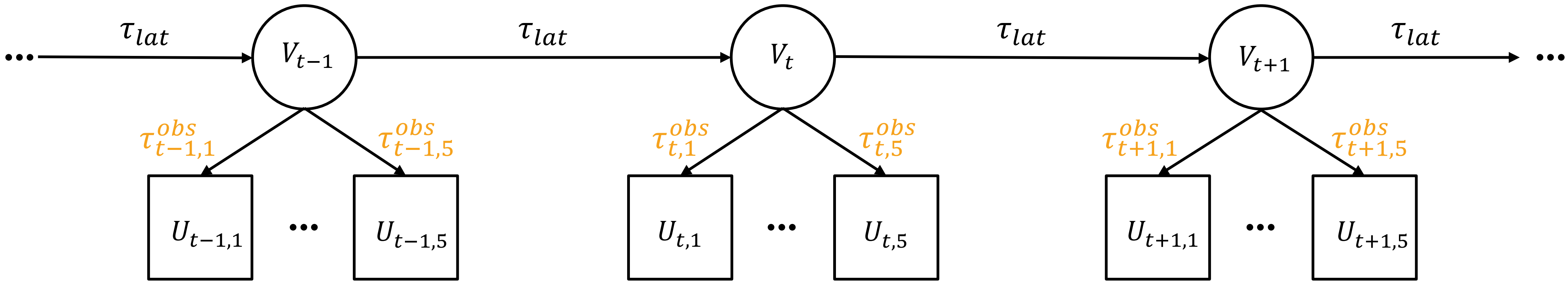


- For $t = 1, \dots, T, j = 1, \dots, 5$:

Observation Equation: $U_{t,j} | V_t = v_t, \tau^{obs}_{t,j}, m^{obs}_j \sim C^{m^{obs}_j}_{U_j|V}(\cdot | v_t; \tau^{obs}_{t,j}),$

State Equation: $V_t | V_{t-1} = v_{t-1}, \tau_{lat}, m_{lat} \sim C^{m_{lat}}_{V_2|V_1}(\cdot | v_{t-1}; \tau_{lat})$

COPULA-SSM WITH TIME-VARYING PARAMETERS



- For $t = 1, \dots, T, j = 1, \dots, 5$:

  Observation Equation: $U_{t,j} | V_t = v_t, \tau_{t,j}^{obs}, m_j^{obs} \sim C_{U_j|V}^{m_j^{obs}}(\cdot | v_t; \tau_{t,j}^{obs}),$

  State Equation: $V_t | V_{t-1} = v_{t-1}, \tau_{lat}, m_{lat} \sim C_{V_2|V_1}^{m_{lat}}(\cdot | v_{t-1}; \tau_{lat})$

- $FZ(\tau_{t,j}^{obs}) = \alpha_j + \beta_j \cdot NCD_t$ → β_j describes dependence of $\tau_{t,j}^{obs}$ on COVID-19 deaths
- $m_{lat}, m_j^{obs} \in \mathcal{M} = \{Gaussian,\ Student\ t,\ Gumbel,\ Clayton\}$

- Latent variable: Captures the unobserved health and lifestyle of the US population

OUTLINE

Data Description

Introduction to Copulas

Marginal Model

Copula Model

Results of Bayesian Inference

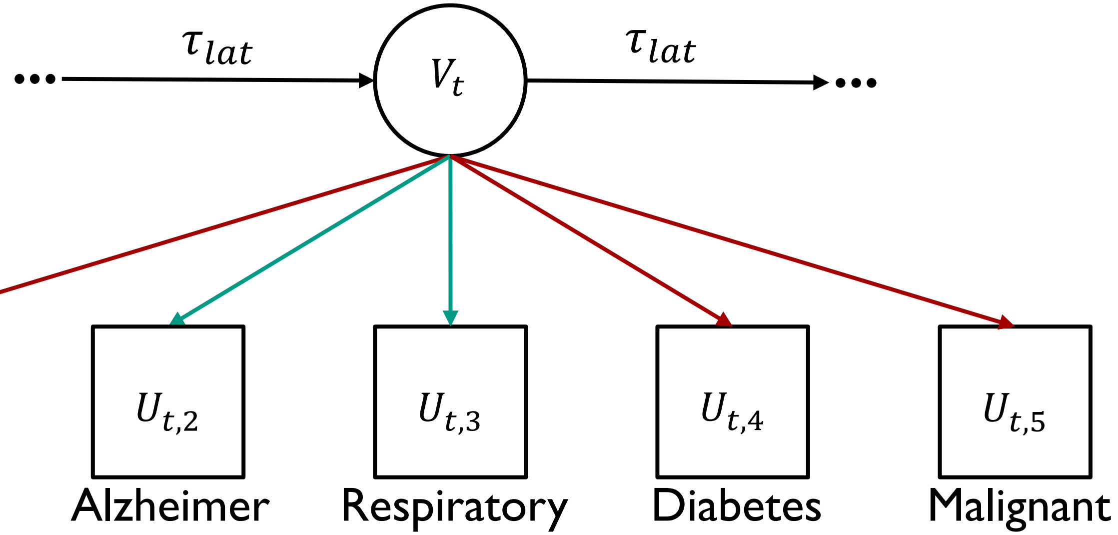
Scenarios

Summary and Outlook

THE MODEL TWODEP

$FZ(\tau_{t,j}^{obs}) = \alpha_j + \beta_j \cdot NCD_t$

$FZ(\tau_{t,j}^{obs}) = \alpha_j$



$FZ(\tau_{t,j}^{obs}) = \alpha_j + \beta_j \cdot NCD_t$

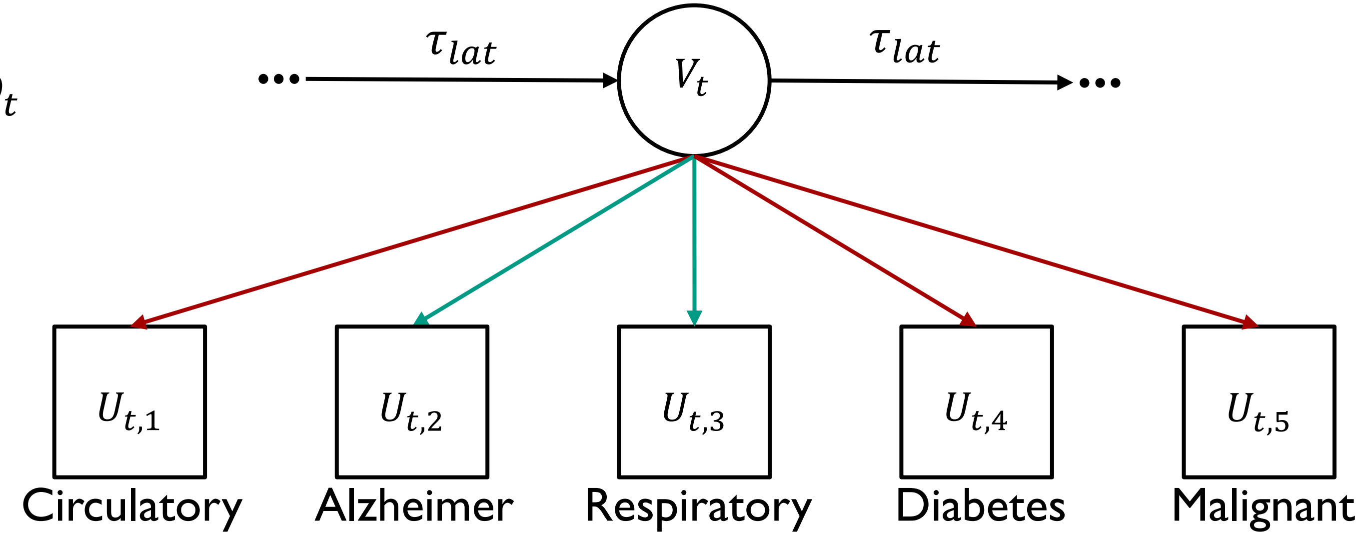
The model:

- Only the dependence between Alzheimer/Respiratory and latent states changes with COVID-19
- Outperforms (1) the Gaussian model, (2) the model where COVID-19 is not taken into account ($\beta_j = 0$), (3) the model where all $\beta_j \neq 0$

THE MODEL TWODEP

$FZ(\tau_{t,j}^{obs}) = \alpha_j + \beta_j \cdot NCD_t$

$FZ(\tau_{t,j}^{obs}) = \alpha_j$



$FZ(\tau_{t,j}^{obs}) = \alpha_j + \beta_j \cdot NCD_t$

Interpretation:

- COVID-19 had a strong influence on the elderly population
- Respiratory and especially Alzheimer are along the leading causes of death in older population
- Both respiratory diseases and COVID-19 directly impact the respiratory system

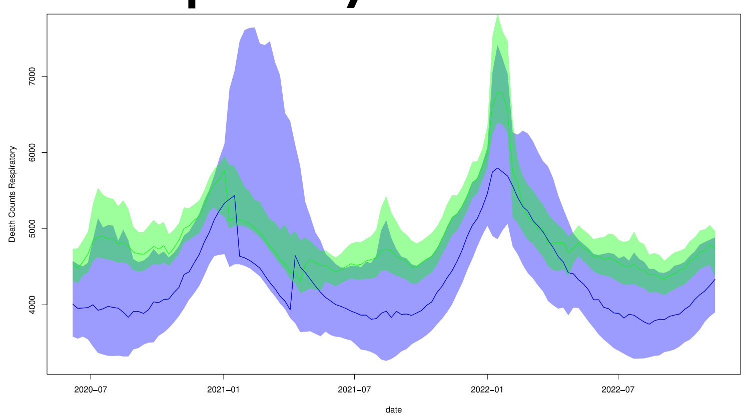
OUTLINE

SETUP OF SIMULATION

- Study pandemic period from June 2020 until November 2022
- What is the effect of a constant low/high number of COVID-19 deaths
- Study two scenarios
- Low-COVID-19 scenario:
  - Set $CD_{low} = 2353$ to the 25% sample quantile of observed values
- High-COVID-19 scenario:
  - Set $CD_{high} = 26026$ to maximum observed value

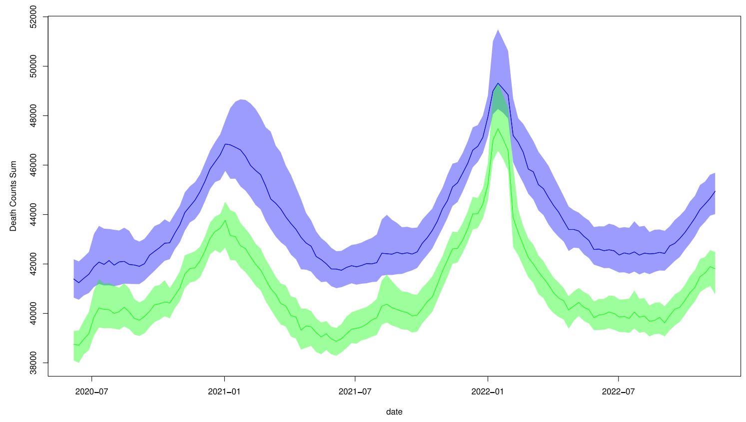
SCENARIO-BASED SIMULATION

Respiratory death counts



Alzheimer death counts

Total death counts



Green: low-COVID-19, Blue: high-COVID-19

- High scenario lowers Respiratory death counts (Competing causes of death)
- High scenario elevates Alzheimer death counts until 2022
- At the end: High scenario lowers Alzheimer (Potentially health policy interventions)
- Overall mortality increases from low to high scenario
- Wider interval for high scenario: Indicates chance of extreme mortality risk

CONCLUSIONS

Contributions

- Utilize copula SSMs to quantify the evolving joint dynamics of 5 causes of death
- Extend copula SSMs by introducing covariates into dependence measure

  → Allows for time-varying dynamics between causes of death

  → Improves flexibility, makes it more realistic

  → Enables us to assess different hypothetical scenarios

Some Findings

- Our findings underscore the pandemic‘s influence on the elderly population
- We find a competing relationship between respiratory diseases and COVID-19

Outlook

- Consider potential impacts across various demographic and socio-economic groups
- Higher-Dimensional Latent Variables

Thank You!